PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/30, G01N 33/68
C07K 15/28, 13/00, C12Q 1/68

A1

(11) International Publication Number: WO 94/04681

(43) International Publication Date: 3 March 1994 (03.03.94)

(21) International Application Number: PCT/IT93/00089

(22) International Filing Date: 10 August 1993 (10.08.93)

(30) Priority data:
RM92000603 13 August 1992 (13.08.92) IT

(71) Applicants *(for all designated States except US)*: ISTITUTO SUPERIORE DI SANITA' [IT/IT]; Viale Regina Elena, 299, I-00161 Roma (IT). UNIVERSITA' DEGLI STUDI DI ROMA "LA SAPIENZA" [IT/IT]; P.le Aldo Moro, 5, I-00185 Roma (IT).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : LA ROSA, Giuseppe [IT/IT]; GOMEZ MORALES, Maria Angeles [ES/IT]; POZIO, Edoardo [IT/IT]; Istituto Superiore di Sanità, Viale Regina Elena, 299, I-00161 Roma (IT). CRISANTI, Andrea [IT/IT]; MULLER, Hans, Michael [DE/IT]; RANUCCI, Lorella [IT/IT]; RECKMANN, Ingeborg [DE/IT]; Università degli Studi di Roma "La Sapienza", P.le Aldo Moro, 5, I-00185 Roma (IT). COLYZZI, Mario [IT/IT]; Università degli Studio di Roma "La Sapienza", P.le Aldo Moro, 5, I-00185 Roma (IT).

(74) Agents: BANCHETTI, Marina et al.; Ing. Barzanò & Zanardo Roma S.p.A., Via Piemonte, 26, I-00187 Roma (IT).

(81) Designated States: AT, AU, BB, BG, BR, BY, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, KZ, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, US, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NUCLEOTIDE SEQUENCES CODING FOR CRYPTOSPORIDIUM PROTEINS, POLYPEPTIDES CODED BY SAID SEQUENCES AND KITS FOR THE USE THEREOF

(57) Abstract

Nucleic acids and polypeptides coded by said nucleic acids, derived from protozoan parasites of *Cryptosporidium* genus are described. Nucleic acids and peptides are advantageously used for developing detection assays of *Cryptosporidium* in biological samples of human and animal origin and/or in the environment.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
LV	Latvia
MC	Monaco
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SI	Slovenia
SK	Slovak Republic
SN	Senegal
TD	Chad
TG	Togo
UA	Ukraine
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

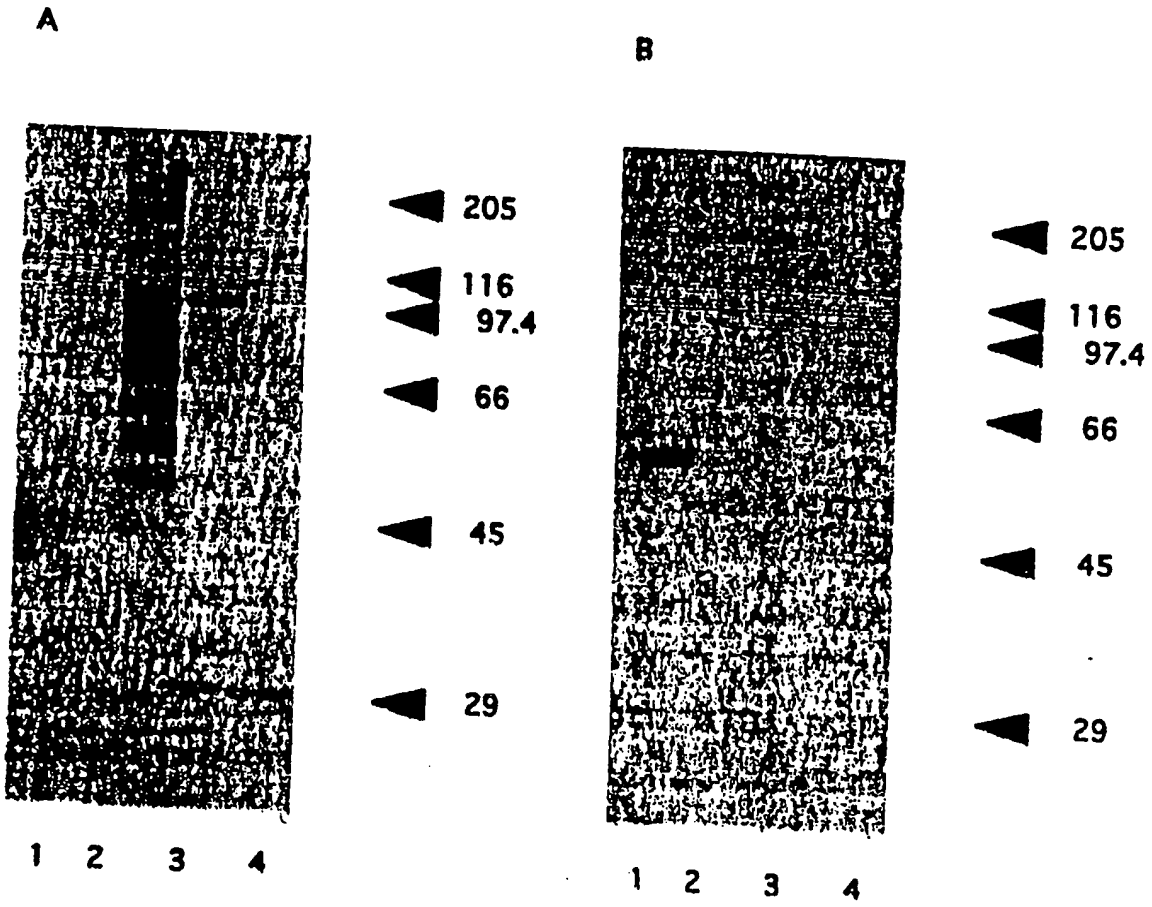


FIG. 1

FIG. 2

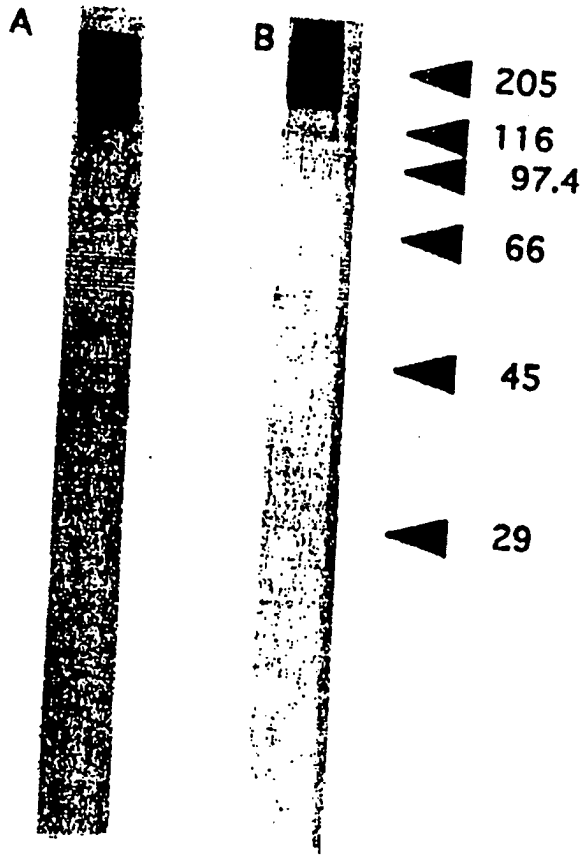
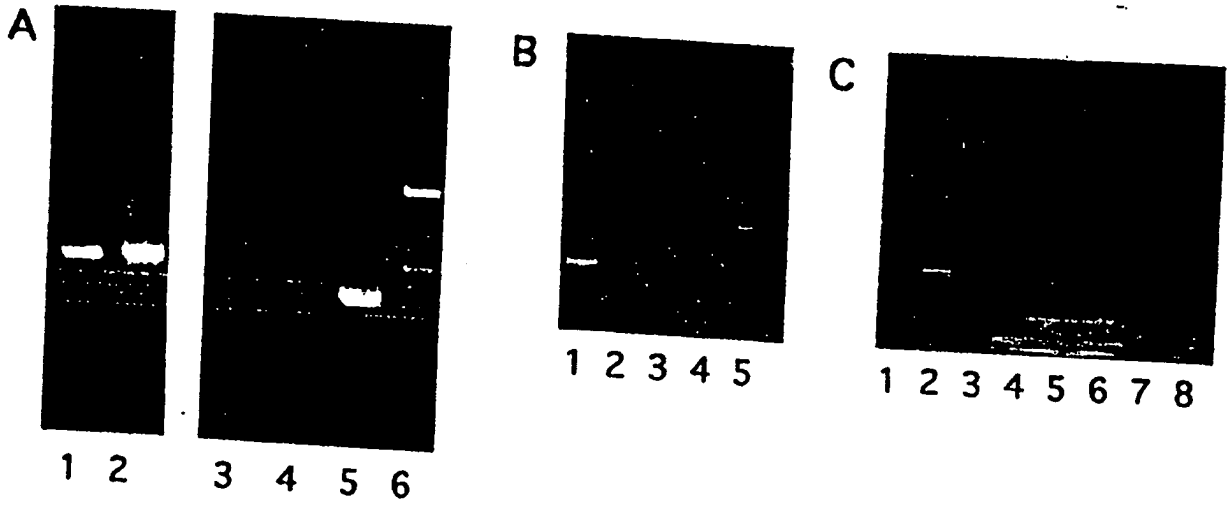


FIG. 3



NUCLEOTIDE SEQUENCES CODING FOR CRYPTOSPORIDIUM PROTEINS, POLYPEPTIDES CODED BY SAID SEQUENCES AND KITS FOR THE USE THEREOF

DESCRIPTION

The invention concerns nucleic acids and polypeptides coded by said nucleic acids, derived from protozoan parasites of *Cryptosporidium* genus. Nucleic acids and peptides are advantageously used for developing detection assays of *Cryptosporidium* in biological samples of human and animal origin and/or in the environment.

Cryptosporidium parasites infect the intestinal tract of several animal species. Over the last decade the number of infections in humans has dramatically increased. Most of the affected patients show a marked immunodeficiency, with a high incidence of AIDS. The immunocompromised patients develop a severe and frequently irreversible diarrhoea which causes malnutrition and represents a major factor leading to death.

The prophylaxis of *Cryptosporidium* infections is hampered by the lack of reliable immunoassays for the detection of the parasite. Moreover, by microscopic examination, parasite oocyst are difficult to discriminate from several microoganisms that are morphologically similar to *Cryptosporidium*, like *Candida* species.

The development of immunological and molecular diagnostic assays highly specific for *Cryptosporidium* absolutely requires the biochemical characterisation and localization of parasite antigens and the cloning of corresponding genes. The availability of diagnostic assays would permit to develop prophylactic measures for immunodeficient patients by detecting the parasites

in the environment (food, water) and in individuals (relatives, nurses) that may act as carriers.

Moreover, no effective therapeutic compounds against *Cryptosporidium* infection are available; therefore there is the need to develop reagents able to prevent the onset of the infection in immunodeficient patients.

The authors of the present invention have obtained a rabbit antiserum raised against a whole lysate of *Cryptosporidium* oocysts. The serum has been used to screen a genomic library of *Cryptosporidium* from infected intestinal mucosal cells, into the expression vector λgt11. Among clones isolated, clone cpRL3 has been shown to comprise a 2359 bp (SEQ ID No.1) insert with an open reading frame coding for a polypeptide of 786 amino acids (SEQ ID No.2). Scanning of the 67.0 version of the "GENE BANK" data base using the cpRL3 sequence failed to reveal any similarity with known DNA sequences.

The cpRL3 insert has been subcloned into an expression plasmid and the corresponding recombinant polypeptide is produced in *E. coli*, fused at the N-terminus to a stretch of six histidines. The histidine sequence allows a fast and efficient purification by nickel chelate chromatography of the polypeptide encoded by cpRL3. The recombinant polypeptide is used as immunogen in Balb/c mice for producing antisera and monoclonal antibodies. The recombinant protein is highly immunogenic. This sequence codes for a portion of a protein of the oocyst wall of Cryptosporidium that has an apparent molecular weight of 190,000 Dalton, named "Cryptosporidium Oocyst Wall Protein" (COWP).

Nucleotide and polypeptide sequences derived from the isolated sequence are advantageously utilized for diagnosis of *Cryptosporidium* infections in patients

and/or animals and for analysis of environmental contamination by *Cryptosporidium* oocysts. Treatment of *Cryptosporidium* infection may be obtained by administration of antibodies raised against COWP.

The invention concerns also: - designing of specific oligonucleotides, as primers in polymerase chain reactions (PCR) reactions for the detection of the COWP DNA sequence; - the utilisation of COWP amino acid sequence or fragments thereof to raise antisera and/or monoclonal antibodies for the development of immunological assays to detect the presence of *Cryptosporidium* and/or COWP molecule; - the use of COWP derived polypeptides, either synthetic or recombinant, as components of diagnostic kits for the detection of COWP released by parasites; - the utilisation of COWP-derived polypeptides, either synthetic or recombinant, for producing antisera and/or monoclonal antibodies to be employed in the therapy of *Cryptosporidiosis*.

It is a specific object of the present invention a polypeptide in a substantially purified form comprising a contiguous sequence coded by a *Cryptosporidium* gene, said gene comprising a nucleotide sequence at least 50 % homologous to the sequence of SEQ ID No.1. Preferably said contiguous sequence comprises an antigenic determinant of *Cryptosporidium*. More preferably said contiguous sequence is coded by the sequence of SEQ ID No.1, or parts thereof; most preferably said contiguous sequence is comprised in the aminoacid sequence of SEQ ID No.2.

It is another object of the invention a diagnostic kit for the detection of *Cryptosporidium* in biological and environmental samples comprising, as specif ligand, the polypeptide according to the invention.

It is another object of the invention the use of a polypeptide according to the invention for raising antibodies able to detect *Cryptosporidium* infection in biological and environmental samples.

It is another object of the invention an antibody obtained using as immunogen a polypeptide according to the invention.

It is another object of the invention a diagnostic kit for the detection of *Cryptosporidium* in biological and environmental samples comprising, as specif ligand, an antibody able to react with at least one polypeptide according to the invention.

It is another object of the invention an oligonucleotide derived from a *Cryptosporidium* gene, said gene comprising a sequence at least 50 % homologous to the sequence of SEQ ID No.1. Preferably said oligonucleotide has a sequence comprised in the sequence of SEQ ID No.1, or in the complementary sequence of SEQ ID No.1.

It is another object of the invention a diagnostic kit for the detection of *Cryptosporidium* in biological and environmental samples comprising, as specif ligand, the oligonucleotide according to the invention.

It is another object of the invention a PCR kit for the amplification of *Cryptosporidium* DNA comprising, as specif primer, at least one oligonucleotide according to the invention. Preferably said PCR kit comprises two oligonucleotides having nucleotide sequences according to the invention.

The invention will be illustrated in the following examples, by making reference to the following figures, wherein:

- Figure 1 shows an immunoblot analysis of the mouse serum (M10/01) (A) and of a control mouse serum anti TRAP (B) against: the expression product of the

control plasmid pDS56/RBSII-E-6his-TRAP purified on nickel column (1); the expression product of the plasmid pDS56/RBSII-E-6his-cpRL3 purified on nickel column (2); a protein lysate of *E. coli* cells transformed with pDS56/RBSII-E-6his-cpRL3 induced with IPTG (3); and non induced (4); molecular weight standard are indicated.

- Figure 2 shows an immunoblot analysis using the mouse serum M10/01 (A) and the MAb1 IB2 (B) against a protein lysate of *C. parvum* oocysts.

- Figure 3A shows an electrophoresis of PCR products amplified from C. parvum DNA (1; 4) and from DNA of the plasmid pDS56/RBSII-E-6his-cpRL3 (2;5) using primers combinations Cry-3/Cry-6 (1;2) or Cry-5/Cry-6 (3;4;5;). As control, PCR reaction without template DNA (3). Figure 3B shows an electrophoresis of PCR products amplified from DNA of several parasite species using primer combination Cry-3/Cry-6: DNA extracted from: *C. parvum* (1), *Sarcocystis sp.* (2), *Giardia lamblia* (3) and *P. falciparum* (4). Figure 3C shows an electrophoresis of PCR products amplified from DNA of progressively diluted *C. parvum* oocysts using the primer combinations Cry-3/Cry-6: molecular standards (1), 160 oocysts (2), 80 oocysts (3), 40 oocysts (4), 20 oocysts (5) 10 oocysts (6). As control, PCR reaction is done without template DNA (7) or in the presence of *P. falciparum* DNA (8). In panels A, B and C, lanes 6, 5 and 1 respectively, DNA markers are 3611; 1166; 606; 517; 396; 318; 263 bp.

Example 1 <u>λgt11 library with DNA extracted from *C. parvum* infected calf intestinal mucosa</u>

To develop a *C. parvum* genomic expression library, DNA extracted from the intestinal mucosa of an infected calf is used.

A newborne calf is infected with 6×10^8 oocysts of *C. parvum* MI ISS-1 (Pozio et al. 1992.

Trans. R. Soc. Trop. Med. Hyg. 86:636-638). After 5 days, the gut is opened, cut into segments of 30 cm each and washed in phosphate-buffered saline (PBS). Nitro-cellulose filters (soaked in PBS) of the same size as the gut segments are applied to the mucosal side for a few seconds. Filters are progressively numbered and processed for DNA extraction. A small sample is removed from each filter and analysed by microscopy to determine whether parasites have been removed. The sample from each filter is incubated with 1% glutaraldehyde in cacodylate buffer for 2 h. The filters are dehydrated through an increasing ethanol series, embedded in Epon, and cured at 60°C for 24 h. Sections are cut at 0.2-μm thickness and stained with toluidine blue. The analysis reveals that nitro-cellulose filters remove only the superficial layer of mucosal cells, together with a large number of parasites.

Only DNA extracted from filters that have removed a large number of parasites are used. DNA is digested with EcoRI and cloned in λgt11 EcoRI digested, 3' end to the coding sequence of β-galactosidase gene. Phage DNA with cloned inserts is packaged *in vitro* (Boehringer Mannheim *in vitro* packaging kit) to generate the library. The quality of the library is evaluated by analysing the sizes of a subset of inserts by polymerase chain reaction (PCR) with oligonucleotides corresponding to the flanking sequences of the EcoRI site of the β-galactosidase gene. The library has a complexity of 4.5×10^6 plaques and an estimated average insert size of 1,800 bp.

The expression library is analysed by use of a rabbit serum developed against purified oocysts of *C. parvum* MI ISS-1. The serum is used after removal of the background reactivity by several absorptions on filters soaked with bacterial and phage lysates. Specific

antibodies bound to filters are detected by use of a second anti-rabbit antibody conjugated to alkaline phosphatase. An insert, named cpRL3, consisting of a 2,359-bp open reading frame encoding a polypeptide of 786 amino acids, is isolated. The lack of both a start codon and a stop codon indicates that the sequence represents part of the coding sequence of the isolated parasitic gene.

Example 2 Expression of the cpRL3 sequence in *E. coli*

The DNA insert cpRL3 is cloned in the EcoRI site of plasmid pDS56/RBSII-E$^-$ 6xHis (a pDS56/RBSII derived plasmid containing an EcoRI site in the polylinker). The expression unit of this vector is under the control of an isopropyl-β-thiogalactopyranoside (IPTG)-inducible promoter and yields a fusion between a stretch of 6 histidines and the amino terminus of the inserted sequences (Stuber et al. 1990, Eur. J. Immunol 20:819-824). cpRL3 is expressed in *E. coli* M15 carrying the lac repressor-producing plasmid pUHA1. Induction is performed in LB medium for 4 h at 37°C; 1 mM IPTG is added when the cell density reaches an optical density at 600 nm of 0 6.

Example 3 Purification of recombinant polypeptide 6x His-cpRL3

The expression product of the cpRL3 sequence (recombinant polypeptide 6xHis-cpRL3) is purified in a single-step procedure, by nickel chelate affinity chromatography (Stuber et al., ibid.). In brief, one litre of an induced culture of M15(pUHA1) cells carrying plasmid pDS56/RBSII-E-6xHis-cpRL3 is harvested and stirred for 3 h in 100 ml of 6 M guanidine hydrochloride, 100 mM Na_2HPO_4, pH 8. The suspension is centrifuged at 10,000 x g, and the supernatant is directly applied to a nickel column (NTA-resin, Diagen). After an equilibration step with 8 M urea, 100

mM NaH_2PO_4, 10 mM Tris, pH 8, 6xHis-cpRL3 is eluted by lowering the pH of the urea solution stepwise to pH 4. From one litre of culture, 1 mg of 6xHis-cpRL3 is obtained.

Example 4 Immunoblotting

Parasite lysates are obtained from *Cryptosporidium* oocysts purified by Percoll gradient centrifugation (Waldman, E., et al. 1986. J. Clin. Microbiol. 23:199-200). Parasites are lysed by incubation of pellet of 2 x 10^7 oocysts with 0.2 ml of sample buffer (33 mM Tris-HCl, pH 6.8, 190 mM glycerol, 0.1% SDS). Bacterial lysates are obtained by treatment of 10^9 induced or non induced *E. coli* cells with 1 ml of sample buffer. Proteins in total cell lysates are separated by SDS-polyacrylamide gel electrophoresis (PAGE) (Laemmli,U.K. 1970. Nature (London) 227:680-685) and electroblotted onto nitro-cellulose filters (blotting buffer, 25 mM Tris, 192 mM glycine, 20% methanol). Non-specific adsorption of antibodies to the nitro-cellulose is prevented by saturation of the filters with 1% bovine serum albumin in 2x TBST (20 mM Tris-HCl, pH 8, 300 mM NaCl, 0.1% Tween 20) for 2 h at room temperature. Nitrocellulose filters are incubated with antibodies for 2 h at room temperature. After extensive washing with 2x TBST, antibodies bound to the filters are detected by use of goat anti-mouse immunoglobulin (heavy and light chains) conjugated to alkaline phosphates (Promega). Phosphatase activity is disclosed by incubation of the filters with 0.3 mg of Nitro Blue Tetrazolium and 0.15 mg of 5-bromo 4-chloro-3-indolyl phosphate per ml in 100 mM Tris-HCl (pH 9.5)-100 mM NaCl-5 mM $MgCl_2$.

Example 5 Monoclonal antibody production

After purification by nickel chelate chromatography, recombinant polypeptide 6xHis-cpRL3 is used as immunogen to develop specific antisera and

monoclonal antibodies. BALB/c mice are immunised three times with 50 μg of purified 6xHis-cpRL3 polypeptide in complete (for the first immunisation) or incomplete Freund's adjuvant. Five days after the last immunisation, mouse spleen cells are fused with x63 Ag 8653 myeloma cells and subsequently screened for antibody production (Kohler,G. and C. Milstein 1975 Nature (London) 256:495-497).

The supernatants of cultures from growing hybrids are tested in an enzyme-linked immunosorbent assay (ELISA) against 6xHisCpRL3. Immunised mice develop an antibody titer of 1/500,000. Figure 1 (A and B) shows the ability of one of the antisera to recognise specifically by immunoblot the recombinant polypeptide encoded by cpRL3. The antiserum recognises the recombinant polypeptide in a total lysate of bacteria in which the expression of the cpRL3 insert is induced with IPTG (Fig. 1A, lane 3). The antiserum also recognize the polypeptide after purification by the nickel chelate chromatography. The specificity of the reaction is demonstrated by the lack of reactivity against *E. coli* proteins, as well as against an unrelated recombinant protein, TRAP, expressed from the control plasmid pDS56/RBSII-6xHis(TRAP), also sharing the six histidine amino-terminal tail. Antibodies developed against the expression product of insert cpRL3 (mouse antiserum and monoclonal antibody 2B11) detect, in the lysate of *Cryptosporidium* oocysts, a protein of the apparent molecular weight of 190,000 Dalton, (Figure 2 A and B). These results indicates that the expression product of the insert cpRL3 is part of a *Clyptosporidium* protein expressed in the parasite oocysts.

Example 6 Immunofluorescence microscopy

Purified parasite oocysts are air dried on a coverslip and fixed in cold acetone for 5 min. Non-

specific binding is prevented by pre-incubation of the samples in PBS containing 1% bovine albumin. Primary antibodies (culture supernatant) are allowed to react for 40 min at room temperature, and the secondary fluorescinated antibody (Becton Dickinson goat anti-mouse) is allowed to react for 20 min. Observation of the samples is carried out with confocal microscopic apparatus (Bio-Rad Laboratories).

The protein is localised by immunofluorescence using the monoclonal antibody 2B11 on a preparation of oocysts fixed with acetone. The monoclonal 2B11 specifically binds to a protein of oocyst wall of *Cryptosporidium*. The result of the immunolocalization is confirmed by confocal immunofluorescence analysis which shows that the reactivity is detected on the surface of the oocyst wall. On the basis of these observations, the Cryptosporidium protein encoded by the insert cpRL3 is named as Cryptosporidium oocyst wall protein, COWP.

By immunofluorescence, monoclonal antibodies can identify as few as 100 oocysts/ml of stools. The detection limit of this method is determined resuspending known numbers of oocysts in samples of non infected stools.

Example 7 ELISA assays

Monoclonal antibodies produced against the product of the insert cpRL3 are able to recognise COWP in solution when used in ELISA assays based on antigen capture with two antibodies. Therefore specific antibodies directed against this protein or fragments thereof having an antigenic activity, as the product of cpRL3, can be employed to develop diagnostic assay to reveal *Cryptosporidium* infection.

Example 8 PCR

For all PCR experiments, the cpRL3 sequence is amplified in a standard 50 µl PCR reaction mixture

(Saiki, R K., et al. 1988. Science 239:487-491; Scharf, S. J. et al. 1986. Science 233:1076-1078) for 35 cycles at 94,5°C (1 min), 58°C (30 sec); 72°C(1 min), with a Lab Line thermal cycler. The final concentration of $MgCl_2$ is 2 mM. Primers used are:
Cry3 (5'GTCCTACTGGATTCACTCTAC-3')
coding strand, nt.722-742 of SEQ ID No. 1;
Cry5 (5'-CCAGGACATCATCATGGTCATTCTCATGGGC-3')
coding strand, nt. 1099-1129 of SEQ ID No. 1;
Cry6 (5'-CCGAATATGTAACACATTTATCCGC-3')
non coding strand, nt. 1828-1852 of complementary strand of SEQ ID No. 1.

For amplification of the cpRL3 sequence from Cryptosporidium oocysts, samples are incubated for 5 min under reducing conditions and boiled for 10 min thereafter. TaqI polymerase is purchased from Perkin-Elmer Co.

The ability of two oligonucleotide combinations to amplify by PCR the sequence of cpRL3 is shown in Figure 3A. Both of oligonucleotide combinations Cry3/Cry6 and Cry5/Cry6 are able to amplify DNA segments of the expected molecular weight from the plasmid pDS56/RBSII-E-6xHis (cpRL3) and from DNA of *Cryptosporidium*. The amplification reaction using oligonucleotides Cry3/Cry6 is highly specific, (figure 3B). Oligonucleotides in fact do not amplify any DNA segment when DNA of other protozoa (*Giardia lamblia, Plasmodium falciparum o Sarcocystis suis hominis*) is used as template. In PCR experiments the oligonucleotide combination Cry3/Cry6 is able to amplify a specific DNA segment from as few as 40 oocysts of *Cryptosporidium*, (Figure 3C). These data indicate that by using combination of oligonucleotides corresponding different regions of cpRL3 sequence, PCR can be employed to detect the presence of *Cryptosporidium*.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Universita' degli Studi di Roma "La Sapienza"
(B) STREET: P.le Aldo Moro 5
(C) CITY: Rome
(E) COUNTRY: Italy
(F) POSTAL CODE (ZIP): 00185

(A) NAME: Istituto Superiore di Sanita'
(B) STREET: V.le Regina Elena 299
(C) CITY: Rome
(E) COUNTRY: Italy
(F) POSTAL CODE (ZIP): 00161

(ii) TITLE OF INVENTION: Nucleotide sequences coding for Cryptosporidium proteins, polypeptides coded by said sequences and kits

(iii) NUMBER OF SEQUENCES: 2

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2359 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Cryptosporidium parvum

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..2359

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
GAA TTC GAA TGC CCA CCA GGT ACA ATT TTA AAA GAT GAT CAA TGT CAA   48
Glu Phe Glu Cys Pro Pro Gly Thr Ile Leu Lys Asp Asp Gln Cys Gln
 1               5                  10                  15

TCG ATA GAA AGA GTT GAT ACA ATT TGT CCA CCA GGG TTT GTA GAT AAT   96
Ser Ile Glu Arg Val Asp Thr Ile Cys Pro Pro Gly Phe Val Asp Asn
            20                  25                  30

GGC GAA GAT TGT GTC CAA TTT TCT GCA CCA GAG AAA ATT TGC CCC CAA  144
Gly Glu Asp Cys Val Gln Phe Ser Ala Pro Glu Lys Ile Cys Pro Gln
        35                  40                  45

GGA TTT TCT CTT TCC GGA AAA CAA TGT GTT AAA ACA GAA TCT GCT CCA  192
Gly Phe Ser Leu Ser Gly Lys Gln Cys Val Lys Thr Glu Ser Ala Pro
    50                  55                  60

AGA TTA ACA GAA TGC CCA CCA GGT ACA ACC TTG GAA AAT AAC AGT TGT  240
Arg Leu Thr Glu Cys Pro Pro Gly Thr Thr Leu Glu Asn Asn Ser Cys
65                  70                  75                  80

ATT TCA TAT GAA CTA GAA GAT GCC ATT TGT CCA CCT GGA TAT CTC GAC  288
Ile Ser Tyr Glu Leu Glu Asp Ala Ile Cys Pro Pro Gly Tyr Leu Asp
                85                  90                  95

AAT GGA TCA GAC TGT GTT CAG TTT TCT CAA CCA GAA AAG GAG TGT CCA  336
Asn Gly Ser Asp Cys Val Gln Phe Ser Gln Pro Glu Lys Glu Cys Pro
            100                 105                 110

ACA GGT TTT GTA TTA ATT GGA AAA CAA TGT ACC CAA ACT ACT CAA GCT  384
Thr Gly Phe Val Leu Ile Gly Lys Gln Cys Thr Gln Thr Thr Gln Ala
        115                 120                 125

CCA CCA CAA CCA GAG TGT CCT CCA GGT ACA AAC CTG GTA AAT GGA CAA  432
Pro Pro Gln Pro Glu Cys Pro Pro Gly Thr Asn Leu Val Asn Gly Gln
    130                 135                 140

TGC CAA AAA GTT GAA AGG ATA AAT ATG GTA TGT CCA ACT GGT TTT ATT  480
Cys Gln Lys Val Glu Arg Ile Asn Met Val Cys Pro Thr Gly Phe Ile
145                 150                 155                 160

GAT AAT GGT ACA AAT TGT GCT TCT TTC TCC GCA CCA AAC AGA GAA TGC  528
Asp Asn Gly Thr Asn Cys Ala Ser Phe Ser Ala Pro Asn Arg Glu Cys
                165                 170                 175

CCA CCT GGA TAT ACA CTT TCT GGA TCC CAA TGC GAG CAA ATA AAA GAA  576
Pro Pro Gly Tyr Thr Leu Ser Gly Ser Gln Cys Glu Gln Ile Lys Glu
            180                 185                 190
```

```
GCA CCT CCT GTT TCA GAA TGT CCA CCA GGA TAT AAA CTT CAA GGA AAT  624
Ala Pro Pro Val Ser Glu Cys Pro Pro Gly Tyr Lys Leu Gln Gly Asn
        195                 200                 205

CAA TGT ACT GCA CTA AAA ATG ATC GAT GCT ATC TGC CCA GAT GGA TTT  672
Gln Cys Thr Ala Leu Lys Met Ile Asp Ala Ile Cys Pro Asp Gly Phe
    210                 215                 220

TTA CCA AAT GGA GAC GAT TGT ATC CAA TTT TCT CCT GCT TCA ACT GTA  720
Leu Pro Asn Gly Asp Asp Cys Ile Gln Phe Ser Pro Ala Ser Thr Val
225                 230                 235                 240

TGT CCT ACT GGA TTC ACT CTA CAA AAT CAA CAG TGT GTT CAA ACA ACT  768
Cys Pro Thr Gly Phe Thr Leu Gln Asn Gln Gln Cys Val Gln Thr Thr
                245                 250                 255

ACC TCA CCA AAA ACA CCA GAA TGT CCT CCA GGT TCT GCG TTG GAT GGA  816
Thr Ser Pro Lys Thr Pro Glu Cys Pro Pro Gly Ser Ala Leu Asp Gly
            260                 265                 270

GAC TCG TGC ACA AGA CTT GTT CCC GGG GCT CTT CAA TAC GTT TGT CCT  864
Asp Ser Cys Thr Arg Leu Val Pro Gly Ala Leu Gln Tyr Val Cys Pro
        275                 280                 285

GTT GGT ACT AGA GAG GGG GAC GTT TGC GTA GAG AGA TCG ATT AGT TCG  912
Val Gly Thr Arg Glu Gly Asp Val Cys Val Glu Arg Ser Ile Ser Ser
    290                 295                 300

CCT GTT TTG GAA TGC CCA CCT GGT TAT TCA TTG GAA ACA GGT AAA CAA  960
Pro Val Leu Glu Cys Pro Pro Gly Tyr Ser Leu Glu Thr Gly Lys Gln
305                 310                 315                 320

TGT GTT AGA AGA AGC CAA TAT GAC TGT TCA GTA ACA ACT TAT GTT ACA 1008
Cys Val Arg Arg Ser Gln Tyr Asp Cys Ser Val Thr Thr Tyr Val Thr
                325                 330                 335

GAG TGT AAA ACA CCT GAT GTT AAA GCA CTA AGA AGA TTA GCA GCT GCA 1056
Glu Cys Lys Thr Pro Asp Val Lys Ala Leu Arg Arg Leu Ala Ala Ala
            340                 345                 350

AAA GAA ACA TCA ACA GTT TAT GAA ACA TCT GAG ATA CAA AAT CCA GGA 1104
Lys Glu Thr Ser Thr Val Tyr Glu Thr Ser Glu Ile Gln Asn Pro Gly
        355                 360                 365

CAT CAT CAT GGT CAT TCT CAT GGG CAT TCA CAT TCA CAA GTT ATA CCA 1152
His His His Gly His Ser His Gly His Ser His Ser Gln Val Ile Pro
    370                 375                 380

ATT CAA ACC CAG AAT ATA CAT ACA CAA CAT CAT AAA GAG GCT CCA AGG 1200
Ile Gln Thr Gln Asn Ile His Thr Gln His His Lys Glu Ala Pro Arg
385                 390                 395                 400
```

```
CCA ATT TGT GAA GAT GTT CCA AAA ATT ACC CCA AAA ACT TGT ACA AAA 1248
Pro Ile Cys Glu Asp Val Pro Lys Ile Thr Pro Lys Thr Cys Thr Lys
                405                 410                 415

GCT GAT TCT GTC CCA GCT GTG CCT ATT TGC GAG AAC AAT GCT GAA CTT 1296
Ala Asp Ser Val Pro Ala Val Pro Ile Cys Glu Asn Asn Ala Glu Leu
            420                 425                 430

GTA GGA AAA GAA TGT GTA TTA ACA AAT TAC TAC CCA TTA GAA GCA ATT 1344
Val Gly Lys Glu Cys Val Leu Thr Asn Tyr Tyr Pro Leu Glu Ala Ile
        435                 440                 445

TGT CAA GAT GGA ACA AGA TCA AAA GAG TGT GCT AAG TTT GTA AAA ACT 1392
Cys Gln Asp Gly Thr Arg Ser Lys Glu Cys Ala Lys Phe Val Lys Thr
    450                 455                 460

CCA CCT ACT TTA AAA TGT CCG CCA GGT TCT GTA GAT GTA GGA TCT CAA 1440
Pro Pro Thr Leu Lys Cys Pro Pro Gly Ser Val Asp Val Gly Ser Gln
465                 470                 475                 480

TGT CAA GTT AAC AAA TAT TCA CCA TAT GAT CTT GCA TGC CCT GCA GGA 1488
Cys Gln Val Asn Lys Tyr Ser Pro Tyr Asp Leu Ala Cys Pro Ala Gly
                485                 490                 495

TAT GCA TTG GTT GGA GAC AAA TGC GCT ACC ACA AGA GAA AAA GTT TGC 1536
Tyr Ala Leu Val Gly Asp Lys Cys Ala Thr Thr Arg Glu Lys Val Cys
            500                 505                 510

CCG AAT GAA AGT TGC CAA AGA GTT GTA ACT GCG CCT GTT TCT TTA ACT 1584
Pro Asn Glu Ser Cys Gln Arg Val Val Thr Ala Pro Val Ser Leu Thr
        515                 520                 525

TGT CCC CCT GGA TAT CAC CAA ATA GAT GAA GTT ATG AAT ATT TCT GCT 1632
Cys Pro Pro Gly Tyr His Gln Ile Asp Glu Val Met Asn Ile Ser Ala
    530                 535                 540

CAT CCA CAC CAC AGA CAC TTA GCT GGG GTT CAA TCT ACT TCT CAA AAG 1680
His Pro His His Arg His Leu Ala Gly Val Gln Ser Thr Ser Gln Lys
545                 550                 555                 560

GGA TAT TCT CAT GGA CAT AAA TAT ACT CCT GTA ATT TCT CAG CCA CCA 1728
Gly Tyr Ser His Gly His Lys Tyr Thr Pro Val Ile Ser Gln Pro Pro
                565                 570                 575

CAA CCA GTT CCA GTT GTT GCT CCT ATT CAG CAA ATG AAA TGC ATC CAT 1776
Gln Pro Val Pro Val Val Ala Pro Ile Gln Gln Met Lys Cys Ile His
            580                 585                 590

GCA AAC CAT GCT CCA TAT AAT CTT ATC TGT CCT GTT GGA TCA AGA CTT 1824
Ala Asn His Ala Pro Tyr Asn Leu Ile Cys Pro Val Gly Ser Arg Leu
        595                 600                 605
```

```
GTA GCG GAT AAA TGT GTT ACA TAT TCG GAT AAA ATA TGT CCA AAT GGT 1872
Val Ala Asp Lys Cys Val Thr Tyr Ser Asp Lys Ile Cys Pro Asn Gly
    610                 615                 620

AAT TGC GAG CGT ATA TAT AAT GAG CCT GCT GAA TTA GTA TGC CCT CCA 1920
Asn Cys Glu Arg Ile Tyr Asn Glu Pro Ala Glu Leu Val Cys Pro Pro
625                 630                 635                 640

GGA TTC TCA TCA TCT AAA CCA ATT CAG CCA ATA AGC CAT TCT CAT ATT 1968
Gly Phe Ser Ser Ser Lys Pro Ile Gln Pro Ile Ser His Ser His Ile
                645                 650                 655

AAC CAT CCA AAT GTT TCT GTT CCC GTC CAA CCA CAA ACT ATT AAC CAA 2016
Asn His Pro Asn Val Ser Val Pro Val Gln Pro Gln Thr Ile Asn Gln
            660                 665                 670

CCA CAA GTA ATT CAA CAA AGA CAA GTA AAT TAT CAG CCA CAA GTA ATT 2064
Pro Gln Val Ile Gln Gln Arg Gln Val Asn Tyr Gln Pro Gln Val Ile
        675                 680                 685

CAT CAA ACA CAG GAA ATT TTA ACA ACT TAT CCA ACT CCA GTT TAC CAA 2112
His Gln Thr Gln Glu Ile Leu Thr Thr Tyr Pro Thr Pro Val Tyr Gln
    690                 695                 700

ACC GGC ACA ATT TAT CAA GGA CAT CAT CAT CAT CAT CAT CAT CAT CAC 2160
Thr Gly Thr Ile Tyr Gln Gly His His His His His His His His His
705                 710                 715                 720

AGA AAT CTA GCT TCC CCT GAG TGC ATT AAG ACA ATT TCA GTA CCT TAT 2208
Arg Asn Leu Ala Ser Pro Glu Cys Ile Lys Thr Ile Ser Val Pro Tyr
                725                 730                 735

ATT TTA AAA TGC GAA TCT CCA TTT ATT TTA GAT GGC GAC AAA TGT ATC 2256
Ile Leu Lys Cys Glu Ser Pro Phe Ile Leu Asp Gly Asp Lys Cys Ile
            740                 745                 750

GAA AAA ACA GAA AAA ATT TGT CTA CAA GGT GAC TGC AGA AAA CAA GTC 2304
Glu Lys Thr Glu Lys Ile Cys Leu Gln Gly Asp Cys Arg Lys Gln Val
        755                 760                 765

GTC GTT CCA CCA ACT CTT TCA TGT CCA CAA GGT TAC AGA AAT GCC AAC 2352
Val Val Pro Pro Thr Leu Ser Cys Pro Gln Gly Tyr Arg Asn Ala Asn
    770                 775                 780

GGA ATT C                                                   2359
Gly Ile
785
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 786 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Glu Phe Glu Cys Pro Pro Gly Thr Ile Leu Lys Asp Asp Gln Cys Gln
 1               5                  10                  15

Ser Ile Glu Arg Val Asp Thr Ile Cys Pro Pro Gly Phe Val Asp Asn
            20                  25                  30

Gly Glu Asp Cys Val Gln Phe Ser Ala Pro Glu Lys Ile Cys Pro Gln
        35                  40                  45

Gly Phe Ser Leu Ser Gly Lys Gln Cys Val Lys Thr Glu Ser Ala Pro
    50                  55                  60

Arg Leu Thr Glu Cys Pro Pro Gly Thr Thr Leu Glu Asn Asn Ser Cys
65                  70                  75                  80

Ile Ser Tyr Glu Leu Glu Asp Ala Ile Cys Pro Pro Gly Tyr Leu Asp
                85                  90                  95

Asn Gly Ser Asp Cys Val Gln Phe Ser Gln Pro Glu Lys Glu Cys Pro
            100                 105                 110

Thr Gly Phe Val Leu Ile Gly Lys Gln Cys Thr Gln Thr Thr Gln Ala
        115                 120                 125

Pro Pro Gln Pro Glu Cys Pro Pro Gly Thr Asn Leu Val Asn Gly Gln
    130                 135                 140

Cys Gln Lys Val Glu Arg Ile Asn Met Val Cys Pro Thr Gly Phe Ile
145                 150                 155                 160

Asp Asn Gly Thr Asn Cys Ala Ser Phe Ser Ala Pro Asn Arg Glu Cys
                165                 170                 175

Pro Pro Gly Tyr Thr Leu Ser Gly Ser Gln Cys Glu Gln Ile Lys Glu
            180                 185                 190

Ala Pro Pro Val Ser Glu Cys Pro Pro Gly Tyr Lys Leu Gln Gly Asn
        195                 200                 205
```

Gln Cys Thr Ala Leu Lys Met Ile Asp Ala Ile Cys Pro Asp Gly Phe
210 215 220

Leu Pro Asn Gly Asp Asp Cys Ile Gln Phe Ser Pro Ala Ser Thr Val
225 230 235 240

Cys Pro Thr Gly Phe Thr Leu Gln Asn Gln Gln Cys Val Gln Thr Thr
245 250 255

Thr Ser Pro Lys Thr Pro Glu Cys Pro Pro Gly Ser Ala Leu Asp Gly
260 265 270

Asp Ser Cys Thr Arg Leu Val Pro Gly Ala Leu Gln Tyr Val Cys Pro
275 280 285

Val Gly Thr Arg Glu Gly Asp Val Cys Val Glu Arg Ser Ile Ser Ser
290 295 300

Pro Val Leu Glu Cys Pro Pro Gly Tyr Ser Leu Glu Thr Gly Lys Gln
305 310 315 320

Cys Val Arg Arg Ser Gln Tyr Asp Cys Ser Val Thr Thr Tyr Val Thr
325 330 335

Glu Cys Lys Thr Pro Asp Val Lys Ala Leu Arg Arg Leu Ala Ala Ala
340 345 350

Lys Glu Thr Ser Thr Val Tyr Glu Thr Ser Glu Ile Gln Asn Pro Gly
355 360 365

His His His Gly His Ser His Gly His Ser His Ser Gln Val Ile Pro
370 375 380

Ile Gln Thr Gln Asn Ile His Thr Gln His His Lys Glu Ala Pro Arg
385 390 395 400

Pro Ile Cys Glu Asp Val Pro Lys Ile Thr Pro Lys Thr Cys Thr Lys
405 410 415

Ala Asp Ser Val Pro Ala Val Pro Ile Cys Glu Asn Asn Ala Glu Leu
420 425 430

Val Gly Lys Glu Cys Val Leu Thr Asn Tyr Tyr Pro Leu Glu Ala Ile
435 440 445

Cys Gln Asp Gly Thr Arg Ser Lys Glu Cys Ala Lys Phe Val Lys Thr
450 455 460

Pro Pro Thr Leu Lys Cys Pro Pro Gly Ser Val Asp Val Gly Ser Gln
465 470 475 480

Cys Gln Val Asn Lys Tyr Ser Pro Tyr Asp Leu Ala Cys Pro Ala Gly
485 490 495

Tyr Ala Leu Val Gly Asp Lys Cys Ala Thr Thr Arg Glu Lys Val Cys
500 505 510

Pro Asn Glu Ser Cys Gln Arg Val Val Thr Ala Pro Val Ser Leu Thr
515 520 525

Cys Pro Pro Gly Tyr His Gln Ile Asp Glu Val Met Asn Ile Ser Ala
530 535 540

His Pro His His Arg His Leu Ala Gly Val Gln Ser Thr Ser Gln Lys
545 550 555 560

Gly Tyr Ser His Gly His Lys Tyr Thr Pro Val Ile Ser Gln Pro Pro
565 570 575

Gln Pro Val Pro Val Val Ala Pro Ile Gln Gln Met Lys Cys Ile His
580 585 590

Ala Asn His Ala Pro Tyr Asn Leu Ile Cys Pro Val Gly Ser Arg Leu
595 600 605

Val Ala Asp Lys Cys Val Thr Tyr Ser Asp Lys Ile Cys Pro Asn Gly
610 615 620

Asn Cys Glu Arg Ile Tyr Asn Glu Pro Ala Glu Leu Val Cys Pro Pro
625 630 635 640

Gly Phe Ser Ser Ser Lys Pro Ile Gln Pro Ile Ser His Ser His Ile
645 650 655

Asn His Pro Asn Val Ser Val Pro Val Gln Pro Gln Thr Ile Asn Gln
660 665 670

Pro Gln Val Ile Gln Gln Arg Gln Val Asn Tyr Gln Pro Gln Val Ile
675 680 685

His Gln Thr Gln Glu Ile Leu Thr Thr Tyr Pro Thr Pro Val Tyr Gln
690 695 700

Thr Gly Thr Ile Tyr Gln Gly His His His His His His His His His
705 710 715 720

Arg Asn Leu Ala Ser Pro Glu Cys Ile Lys Thr Ile Ser Val Pro Tyr
725 730 735

Ile Leu Lys Cys Glu Ser Pro Phe Ile Leu Asp Gly Asp Lys Cys Ile
740 745 750

Glu Lys Thr Glu Lys Ile Cys Leu Gln Gly Asp Cys Arg Lys Gln Val
755 760 765

Val Val Pro Pro Thr Leu Ser Cys Pro Gln Gly Tyr Arg Asn Ala Asn
770 775 780

Gly Ile
785

CLAIMS

1. A polypeptide in a substantially purified form comprising a contiguous sequence coded by a *Cryptosporidium* gene, said gene comprising a nucleotide sequence at least 50 % homologous to the sequence of SEQ ID No.1.

2. A polypeptide according to Claim 1 wherein said contiguous sequence comprises an antigenic determinant of *Cryptosporidium*.

3. A polypeptide according to any of previous claims wherein said contiguous sequence is coded by the sequence of SEQ ID No.1, or parts thereof.

4. A polypeptide according to Claim 3 wherein said contiguous sequence is comprised in the aminoacid sequence of SEQ ID No.2.

5. A diagnostic kit for the detection of *Cryptosporidium* in biological and environmental samples comprising, as specif ligand, the polypeptide according to the invention.

6. Use of a polypeptide according to the invention for raising antibodies able to detect *Cryptosporidium* infection in biological and environmental samples.

7. An antibody obtained using as immunogen a polypeptide according to any of Claims from 1 to 4.

8. A diagnostic kit for the detection of *Cryptosporidium* in biological and environmental samples comprising, as specif ligand, an antibody able to react with at least one polypeptide according to any of Claims from 1 to 4.

9. An oligonucleotide derived from a *Cryptosporidium* gene, said gene comprising a sequence at least 50 % homologous to the sequence of SEQ ID No.1.

10. An oligonucleotide according to Claim 9 having a sequence comprised in the sequence of SEQ ID No.1, or in the complementary strand of SEQ ID No. 1.

11. A diagnostic kit for the detection of *Cryptosporidium* in biological and environmental samples comprising, as specif ligand, the oligonucleotide according to Claim 9 or 10.

12. A PCR kit for the amplification of *Cryptosporidium* DNA comprising, as specif primer, at least one oligonucleotide according to Claim 9 or 10.

13. A PCR kit according to Claim 12 comprising, as specif primer, two oligonucleotides according to Claim 9 or 10.